

# MuLMINet: Multi-Layer Multi-Input Transformer Network with Weighted Loss

Minwoo Seong<sup>1,‡</sup>, Jeongseok Oh<sup>1,‡</sup> and SeungJun Kim<sup>1\*</sup>

<sup>1</sup> Gwangju Institute of Science and Technology

{seongminwoo, jungseok17}@gm.gist.ac.kr, seungjun@gist.ac.kr

<sup>‡</sup> These authors contributed equally to this work. \* Correspondence Author

## Abstract

The increasing use of artificial intelligence (AI) technology in turn-based sports, such as badminton, has sparked significant interest in evaluating strategies through the analysis of match video data. Predicting future shots based on past ones plays a vital role in coaching and strategic planning. In this study, we present a Multi-Layer Multi-Input Transformer Network (MuLMINet) that leverages professional badminton player match data to accurately predict future shot types and area coordinates. Our approach resulted in achieving the runner-up (2nd place) in the IJCAI CoachAI Badminton Challenge 2023, Track 2. To facilitate further research, we have made our code publicly accessible online, contributing to the broader research community’s knowledge and advancements in the field of AI-assisted sports analysis.

## 1 Introduction

The advent of artificial intelligence is bringing significant changes to sports analytics, with machine learning and deep learning playing key roles in this change [Du and Peng, 2023][Azad *et al.*, 2022]. These techniques are not only employed for evaluating the capabilities and performance of individual players [Hülsmann *et al.*, 2018][Oagaz *et al.*, 2021][Ghosh *et al.*, 2022], but they also play a crucial role in analyzing and assessing the strategies that emerge from interactions between players [Azad *et al.*, 2022][Han *et al.*, 2022][Chen *et al.*, 2022][Chang *et al.*, 2023][Wang *et al.*, 2022b][Wang *et al.*, 2022a]. There have been several attempts to utilize artificial intelligence (AI) technology to evaluate players’ strategies, especially in turn-based sports such as badminton. In particular, the ability to predict future shots based on past ones is integral to player coaching and strategic planning [Wang *et al.*, 2021]. The analysis of shot types that yield successful returns and the identification of optimal ball drop locations can significantly aid players in decision-making processes, thereby enhancing their performance. Several attempts have been made to utilize AI technology to evaluate players’ strategies, especially in turn-based sports such as badminton. Wang *et al.* proposed a language known as Badminton Language from

Shot to Rally (BLSR) to represent the rally process and employed Convolutional Neural Networks (CNN) and encoder-based networks to predict shot influence [Wang *et al.*, 2021][Wang *et al.*, 2022a]. Wang *et al.* also proposed a ShuttleNet network, which predicts the sequence of badminton matches based on a transformer model. In previously proposed models, the prediction of future strokes was primarily based on two variables: shot type and area coordinates. However, these features may not fully capture the complexity of the game as there are numerous factors that influence shot type and area. Recognizing this limitation, we propose a new BLSR prediction network, Multi-Layer Multi-Input Transformer Network (MuLMINet). We introduced our approach in the context of the IJCAI-2023 CoachAI Badminton Challenge Track 2: Forecasting Future Turn-Based Strokes in Badminton Rallies and won second position. This paper delineates our proposed methodology and the corresponding evaluation process, and the code is available on our GitHub repository: <https://github.com/stan5dard/IJCAI-CoachAI-Challenge-2023>.

## 2 Method

### 2.1 Evaluation Metrics of IJCAI CoachAI Badminton Challenge Track 2

In this challenge, the objective is to predict the future stroke shot type and area coordinates based on the data from the first 4 strokes. The evaluation metrics used for this task are defined as follows:

$$Score = \min(l_1, l_2, \dots, l_6) \quad (1)$$

$$l_i = AVG(CE + MAE) \quad (2)$$

The shot type is evaluated using the Cross-Entropy (CE) loss, which measures the discrepancy between the predicted and actual shot-type probabilities. The area coordinate is evaluated using the Mean Absolute Error (MAE), which quantifies the average difference between the predicted and actual area coordinates. To calculate the final score, we take the average of six evaluation metrics and select the minimum value as the final score. This score serves as an overall measure of the predictive model’s performance in predicting stroke shot type and landing coordination. The team with the lowest loss emerges as the winner of this challenge.

## 2.2 Dataset and Data Preprocessing

The competition dataset was based on the ShuttleSet dataset [Du and Peng, 2023] and consisted of three separate files: train.csv, validation.csv, and test.csv. The dataset included a total of 30,172 strokes in the training set, 1,400 strokes in the validation set, and 2,040 strokes in the test set. The training dataset, which included data with a rally length of five or more, formed the basis for model development. In contrast, the validation and test datasets only included data for the first four rallies. Participants were required to construct their own models using the training dataset, and these models were then tested on the validation and test sets as part of the evaluation process. The dataset is divided into four primary categories.

- Rally category: round score A, round score B, get point player, and lose reason
- Temporal category: time number and frame number
- Spatial category: player location area, player location x, player location y, opponent location area, opponent location x, opponent location y, landing area, landing x, landing y, and landing height.
- Hitting category: backhand, around head, and shot type

In order to determine the input features for the model, we examined the correlation between variables. For categorical variables, we calculated the Cramer’s V correlation matrix. Cramer’s V is a measure of association used to assess the strength and direction of the relationship between two categorical variables.

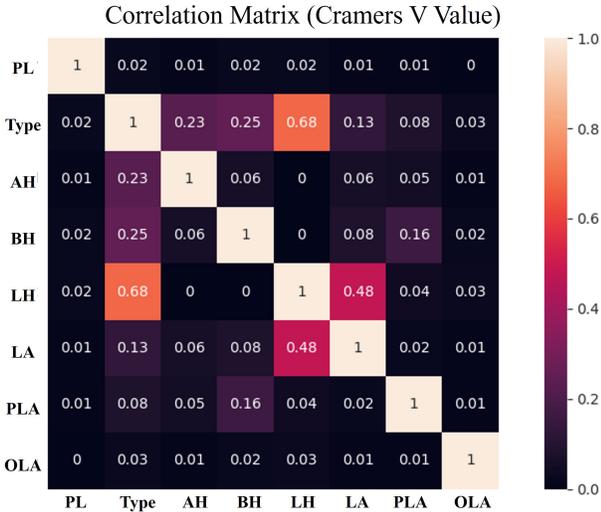


Figure 1: Cramer’s V Correlation Matrix between features: PL denotes player, AH denotes aroundhead, BH denotes backhand, LH denotes landing height, LA denotes landing area, PLA denotes player location area, and OLA denotes opponent location area

Variables such as ”aroundhead” (Cramér’s V = 0.23) and ”backhand” (Cramér’s V = 0.25) demonstrated moderate correlations with shot type. On the other hand, landing height exhibited a strong correlation (Cramér’s V = 0.68) with shot type. Based on these findings, we identified ”aroundhead,”

”backhand,” and landing height as additional input features for our model.

Moreover, for the prediction of area coordinates, we included player area location and opponent area location as additional features. As a result, we utilized a total of eight features as input variables for our predictive model. This selection of features was guided by the observed correlations and their potential relevance in predicting the desired outcomes. By incorporating these variables, we aimed to enhance the model’s accuracy and predictive capabilities.

## 2.3 Network Architecture Design

The network architecture was specifically designed to accommodate additional input features (see Figure 2). We built our network based on the principles established by ShuttleNet, which incorporated player and rally encoders and decoders using BLSR data. We primarily used two encoding methods to process the input data.

Encoded Type 1 involved representing the shot type, aroundhead, backhand, height, player location, opponent location, and player variables within the input vector. By incorporating these variables together, we aimed to capture the relevant characteristics of the shot type and its associated attributes. Encoded Type 2 focused on encoding the area coordinate, aroundhead, backhand, height, player location, opponent location, and player variables. By including these features within the input vector, we aimed to capture the spatial information and positional attributes related to the area coordinates.

This design choice was motivated by the observed correlations between shot type, area coordinate, and other variables. By embedding these variables together, we aimed to achieve improved prediction performance. Similarly, for the decoder output, Type 1 and Type 2 were each passed through the Position Aware Gated Fusion Network. This network was responsible for predicting area, shot type, aroundhead, backhand, height, player location, and opponent location for each observed player.

## 2.4 Loss Function Design and Hyper-parameter tuning

We designed our total loss function as follows:

$$L_T = \alpha(L_{ST} + L_{SL}) + (1 - \alpha)(L_B + L_A + L_H + L_{PL} + L_{OL}) \quad (3)$$

$L_T$  denotes total loss,  $L_{ST}$  denotes loss of shot type,  $L_{SL}$  denotes loss of shot landing coordination,  $L_B$  denotes loss of backhand,  $L_A$  denotes loss of around-head,  $L_H$  denotes loss of landing height,  $L_{PL}$  denotes loss of player location area, and  $L_{OL}$  denotes loss of opponent location area. Alpha ( $\alpha$ ) is a hyperparameter that controls the trade-off between the shot type/location loss and the other losses.

The model was evaluated using a dataset divided into three subsets: train set, validation set, and test set, derived from the train.csv dataset. The test set remained fixed, while the train set and validation set were evaluated using a 5-fold cross-validation approach. During the evaluation process, the model underwent 300 epochs with a fixed learning rate of 0.0001 and a batch size of 32. Hyper-parameter evaluation was conducted by varying the Dimension and Alpha values

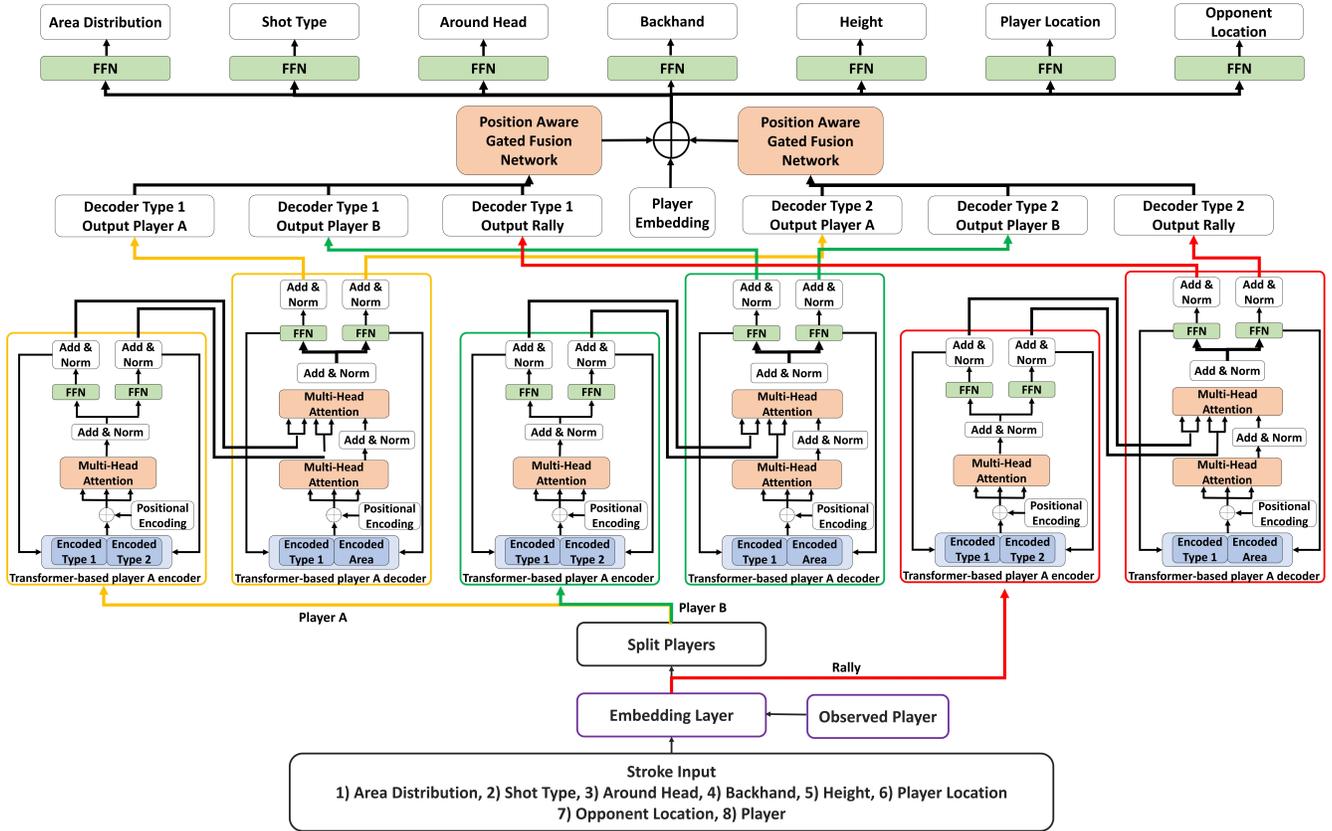


Figure 2: Network Architecture: MuLMINet

(see Table 1). Scores were computed based on the evaluation metric specified in the challenge, and the results were summarized in a table. The average value and standard deviation of the scores were calculated to assess the model’s performance.

Hyper-parameter	Values
Learning rate	0.0001
Batch size	32
Dimension	32, 64, 128
Alpha	0.3, 0.35, 0.4, 0.45
Layer number	1, 2, 3
Epoch	300

Table 1: Hyper-parameter for MuLMINet evaluation

### 3 Results

#### 3.1 Loss Selection Module

To identify the optimal classifier and regressor, we implemented a Loss Selection Module (see Figure 3) that evaluated 72 different hyperparameter cases. The purpose of this module was to determine the ideal combination of models by comparing the average loss across 5-fold cross-validation. For each hyperparameter combination, we assessed the av-

erage loss and identified the combinations that exhibited accurate predictions for shot type and area coordinates. These selected combinations were then submitted to the final challenge as our model entries, reflecting the comprehensive evaluation conducted through the Loss Selection Module.

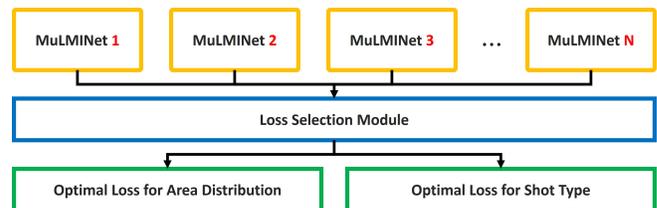


Figure 3: Loss Selection Module

#### 3.2 Optimal Parameter for MuLMINet

After implementing the Loss Selection Module, we successfully identified the best case through a comprehensive evaluation of 36 different hyperparameter combinations. Based on the evaluation results, we selected the combinations that exhibited accurate predictions for shot type and area coordinates. The final combination was submitted as a model entry for the challenge, and Table 2 shows the score from the validation process and the final score of the challenge.

Phase	Total loss	Area loss	Shot loss
Validation phase	2.489	0.6674	1.8216
Testing phase	2.5830	0.7703	1.8127

Table 2: Loss score of validation and testing phase

## 4 Discussions

In the context of our study, it is worth exploring the potential benefits of treating the inputs for area and shot type differently based on their correlations. Although our initial approach involved embedding all features together in Encoded Type 1 and Encoded Type 2, regardless of their associations with shot type and area, we acknowledge that incorporating highly relevant features for each prediction task could yield improvements in accuracy.

For future research, we recommend exploring alternative embedding strategies that take into account feature correlations and tailoring them to the specific requirements of each prediction task. Such investigations could provide insights into optimizing model performance and achieving even lower loss in area coordinates and shot-type prediction tasks.

## 5 Conclusions

In conclusion, we propose MuLMINet, a novel architecture that leverages multi-layer and multi-input approaches to accurately predict both shot types and areas on the BSLR dataset. By integrating a weighted sum loss function and a loss selection module, we derive optimal parameters and models. As a result, we achieve the runner-up position (2nd winner) in the IJCAI CoachAI Badminton Challenge.

## Acknowledgments

This work was supported by the GIST-MIT Research Collaboration grant, which was funded by GIST in 2023. We appreciate the high-performance GPU computing support of HPC-AI Open Infrastructure via GIST SCENT.

## References

- [Azad *et al.*, 2022] Abdus Salam Azad, Edward Kim, Qiancheng Wu, Kimin Lee, Ion Stoica, Pieter Abbeel, Alberto Sangiovanni-Vincentelli, and Sanjit A Seshia. Programmatic modeling and generation of real-time strategic soccer environments for reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6028–6036, 2022.
- [Chang *et al.*, 2023] Kai-Shiang Chang, Wei-Yao Wang, and Wen-Chih Peng. Where will players move next? dynamic graphs and hierarchical fusion for movement forecasting in badminton. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6998–7005, 2023.
- [Chen *et al.*, 2022] Xiusi Chen, Jyun-Yu Jiang, Kun Jin, Yichao Zhou, Mingyan Liu, P Jeffrey Brantingham, and Wei Wang. Reliable: Offline reinforcement learning for tactical strategies in professional basketball games. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 3023–3032, 2022.
- [Du and Peng, 2023] Wei-Yao Wang and Wei-Wei Du and Wen-Chih Peng. ShuttleSet22: Benchmarking stroke forecasting with stroke-level badminton dataset. *CoRR*, abs/2306.15664, 2023.
- [Ghosh *et al.*, 2022] Indrajeet Ghosh, Sreenivasan Ramasamy Ramamurthy, Avijoy Chakma, and Nirmalya Roy. Decoach: Deep learning-based coaching for badminton player assessment. *Pervasive and Mobile Computing*, 83:101608, 2022.
- [Han *et al.*, 2022] Yewon Han, Jaeho Kim, Hon Keung Tony Ng, and Seong W Kim. Logistic regression model for a bivariate binomial distribution with applications in baseball data analysis. *Entropy*, 24(8):1138, 2022.
- [Hülsmann *et al.*, 2018] Felix Hülsmann, Jan Philip Göpfert, Barbara Hammer, Stefan Kopp, and Mario Botsch. Classification of motor errors to provide real-time feedback for sports coaching in virtual reality—a case study in squats and tai chi pushes. *Computers & Graphics*, 76:47–59, 2018.
- [Oagaz *et al.*, 2021] Hawkar Oagaz, Breawn Schoun, and Min-Hyung Choi. Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4332–4343, 2021.
- [Wang *et al.*, 2021] Wei-Yao Wang, Teng-Fong Chan, Hui-Kuo Yang, Chih-Chuan Wang, Yao-Chung Fan, and Wen-Chih Peng. Exploring the long short-term dependencies to infer shot influence in badminton matches. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 1397–1402. IEEE, 2021.
- [Wang *et al.*, 2022a] Wei-Yao Wang, Teng-Fong Chan, Wen-Chih Peng, Hui-Kuo Yang, Chih-Chuan Wang, and Yao-Chung Fan. How is the stroke? inferring shot influence in badminton matches via long short-term dependencies. *ACM Transactions on Intelligent Systems and Technology*, 14(1):1–22, 2022.
- [Wang *et al.*, 2022b] Wei-Yao Wang, Hong-Han Shuai, Kai-Shiang Chang, and Wen-Chih Peng. Shuttlenet: Position-aware fusion of rally progress and player styles for stroke forecasting in badminton. In *AAAI*, pages 4219–4227. AAAI Press, 2022.